

IX Всеукраїнська студентська науково - технічна конференція "ПРИРОДНИЧІ ТА ГУМАНІТАРНІ НАУКИ.
АКТУАЛЬНІ ПИТАННЯ"

УДК 004.73; 004.77

Кіліштоф М. – ст. гр. СНМ-52

Тернопільський національний технічний університет імені Івана Пулюя

СЕРВІСИ РОБОТИ З ГЛИБИННИМ ВЕБОМ

Науковий керівник: ст. викладач Маєвський О.В.

Kilistof M.

Ternopil Ivan Pul'uy National Technical University

SERVICES WORK DEEP WEB

Supervisor: Majevskiy A.

Ключові слова: сервіс, ресурс, «глибинний веб»

Keywords: reception, resource, "Deep web"

Традиційні пошукові системи прагнуть звужити простір глибинного веб, поступово захоплюючи такі ніші, як блоги, наукові сайти, інформаційні агентства. Так, в якості допоміжних сервісів для пошуку по глибинному веб від Google можна рекомендувати: Google Book Search (books.google.com) – пошук книг, Google Scholar (scholar.google.com) – пошук наукових публікацій, Google Code Search (code.google.com) – пошук програмного коду.

Система Goldfire Research від компанії Invention Machine Corp. дозволяє обробляти контент глибинного веб, розміщеного на більш ніж 2000 сайтах урядових, академічних, дослідницьких і комерційних організацій США. Система Goldfire Research володіє інформацією про механізми доступу до баз даних глибинного веб і автоматично генерує запити до них.

Дослідницька пошукова система Infovell з Каліфорнійського університету Берклі дозволяє шукати в глибинному веб по «ключовим фразам», від параграфів до цілих документів, або навіть наборам документів загальним об'ємом до 25 тисяч слів. Система Infovell не залежить від мови, користувачі можуть шукати сторінки англійською, арабською, китайською мовами або ж вводити в рядку пошуку математичні рівняння, хімічні формули.

Існуючі засоби аналізу і просування веб-ресурсів дозволяють по-новому підійти до оцінки співвідношення об'ємів видимого і глибинного веб. Так на веб-сайті www.cu-prg.com приводиться інформація про реальну кількість документів на досліджуваному веб-сайті, представленою в RUNet, і про кількість документів, заіндексовану різними пошуковими системами, в тому числі, Google і Яндекс. Отримавши репрезентативну вибірку по сайтам, наприклад, за рейтингом Рамблера top100, можна отримати оцінку співвідношення видимої та глибинної частини в RUNet-сегменті веб-простору.

Як показують розрахунки, об'єм інформації, що опинилася в глибинній частині веб-простору, перевищує об'єм інформації з видимої частини приблизно в 3-5 разів. Виявляється, за рідкісним виключенням, що чим більший ресурс, тим більша його частина відноситься до глибинного веб. В цьому сенсі невеликі веб-ресурси виграють в доступності. Оскільки велика частка новинних документів виявляється в глибинному веб, то для завдань бізнес-аналітики потрібні спеціальні сервіси доступу до такої інформації. Саме такий сервіс надають служби інтеграції новинному контенту – архіви мережеских ЗМІ. Українські бізнес-аналітики активно використовують найбільші архіви інформації з відкритих джерел «Інтегрум» і InfoStream. Саме використання відкритих

джерел дозволяє конкурентній розвідці діяти в рамках правового поля, але, при цьому, мати високу ефективність.

Можна констатувати, що чим швидше росте веб-простір, тим гірше воно охоплюється традиційними каталогами і пошуковим машинам. Через зростання кількості веб-сайтів і порталів, що використовують бази даних, динамічні системи керування контентом, появи нових версій форматів представлення інформації глибинний веб зростає дуже інтенсивно. З одного боку, Інтернет як величезне сховище збільшує об'єми інформації, доступної «в принципі», але з іншого боку – зростає інформаційний хаос, збільшується ентропія мережевого Інформаційного простору. Все менша частина інформаційних ресурсів стає доступною користувачам реально.

Провідні пошукові системи як і раніше намагаються знайти технічні можливості для індексації вмісту баз даних і отримати доступ до приватних веб-сайтів, проте, їх завдання об'єктивно розходяться із завданнями бізнес-аналітиків – орієнтація традиційних пошукових служб на масовий сервіс в даному випадку виправдана. Таким чином, ніша для систем пошуку в глибинному веб стає все ширшою.

УДК 004.02; 004.6

Ковальська М. – ст. гр. СКмз-61

Тернопільський національний технічний університет імені Івана Пулюя

ПРО ПОНЯТТЯ «ГЛИБИННОГО» ВЕБУ

Науковий керівник: ст. викладач Маєвський О.В.

Koval's'ka M.

Ternopil Ivan Pul'uy National Technical University

ON THE CONCEPT OF "DEEP" WEB

Supervisor: Majevskiy A.

Ключові слова: Веб-простір, «глибинний» веб

Keywords: Web space, "deep" web

Останні дослідження веб-простору показали, що доступні через традиційні інформаційно-пошукові системи більше трильйона веб-сторінок – це лише «видима частина айсберга».

Важливою проблемою є пошук інформації в «прихованому» або «глибинному» веб-просторі, де міститься незрівнянно більша кількість даних, потенціально цікавих для конкурентної розвідки, ніж у відкритій частині Інтернету.

Це, перш за все, динамічні веб-сторінки, інформація з численних баз даних, які можуть представляти великий інтерес для аналітичної роботи. До розряду «прихованого» веб відносяться і повнотекстові інформаційні системи типу LexisNexis або Factiva.

До «прихованих» ресурсів мережі Інтернет можна віднести також пірінгові мережі, такі як BitTorrent, EDonkey, EMule, Gnutella, Kazaa.

Відомо, що необхідної (в тому числі і для конкурентної розвідки) інформації в мережі Інтернет значно більше, ніж її охоплюють універсальні пошукові машини.